

Workshop Summary
International Cooperative for Aerosol Prediction Workshop
On Aerosol Forecast Verification

Angela Benedetti¹, Jeffrey S. Reid², and Peter R. Colarco³

¹European Center for Medium Range Weather Forecasts;

²Naval Research Laboratory;

³NASA Goddard Space Flight Center

Submitted to Bulletin of the American Meteorological Society
February, 2011

Corresponding Author: Angela Benedetti, ECMWF, Shinfield Park, Reading, RG2 9AX,
United Kingdom. Email: Angela.Benedetti@ecmwf.int

Meeting Summary

International Cooperative for Aerosol Prediction Workshop on Aerosol Forecast Verification

BY Angela Benedetti, Jeffrey S. Reid, and Peter R. Colarco

AFFILIATIONS: Benedetti-European Center for Medium Range Weather Forecasts;
Reid -Naval Research Laboratory; Colarco-NASA Goddard Space Flight Center

CORRESPONDING AUTHOR: Angela Benedetti, ECMWF, Shinfield Park, Reading, RG2 9AX, United Kingdom. Email: Angela.Benedetti@ecmwf.int

DATA BOX

TITLE: ICAP Workshop on Aerosol Forecast Verification

What: The purpose of this workshop was to reinforce the working partnership between centers who are actively involved in global aerosol forecasting, and to discuss issues related to forecast verification. Participants included representatives from operational centers with global aerosol forecasting requirements, a panel of experts on Numerical Weather Prediction and Air Quality forecast verification, data providers, and several observers from the research community. The presentations centered on a review of current NWP and AQ practices with subsequent discussion focused on the challenges in defining appropriate verification measures for the next generation of aerosol forecast systems.

When: 30 September-1 October 2010

Where: Oxford, United Kingdom

BEGIN SIDEBAR

General Conclusions and Outcomes of the Meeting:

- 1) An inventory was made of current verification practices at Numerical Weather Prediction and Air Quality Forecasting Centers with the goal of understanding what can be applied directly to the aerosol forecast verification problem. At the same time, the need emerged to define verification measures specific to the needs of the aerosol forecast user community since it was recognized that these measures ultimately inform the direction of the system development and the research efforts. Meeting presentations and summary slides can be downloaded from <http://bobcat.aero.und.edu/jzhang/ICAP>.
- 2) While verification of modeled aerosol fields is a very active area of research among various communities, the specific requirements of operational centers involve the availability of verification data in Near Real Time. This has serious implications on the

effort required from the data providers, which very often operate on low budgets, and are not equipped to support operational activities. Involvement of agencies that deal with operational data delivery is crucial. Benefits of this involvement are foreseen also for the climate modeling community, as data made available for NRT verification, can be reprocessed and made suitable for the needs of that community as well (i.e., transitioning from Level 1/Level 2 to Level 3 products).

3) Collaboration between the various operational centers is deemed essential to establish common scoring practices that would be acceptable to the majority. The input from the data providers is also important in defining these measures, especially when verifying with “unconventional” observations such as aerosol lidar backscatter profiles. Consensus climatologies are needed to provide baseline for comparisons.

4) The success of this type of grassroots organization is based on the open communication between the members of the Cooperative. To this end, there was a general consensus to continue with these informal meetings of the global aerosol forecasting system developers. The next of such meetings will focus on multi-model ensemble forecasting and data assimilation systems for aerosols. Other topics that will be discussed in the near future will include developments in aerosol assimilation and direct aerosol radiance assimilation, to mention a few.

5) A mailing list has been recently set-up with the intent to keep the communication flowing. Please subscribe at <https://lists.nasa.gov/mailman/listinfo/icap-aerosols>, if interested in current and future ICAP activities.

END SIDEBAR

Meeting Summary:

The newly formed International Cooperative on Aerosol Prediction met in Oxford, UK, to discuss specific needs related to global aerosol forecasting verification. While the dynamical meteorology community has a well developed, near real-time observing system to support forecasting and verification activities, the aerosol community is only beginning to address these aspects. Meeting participants included representatives from scientific and operational centers with global aerosol forecasting requirements (ECMWF, FNMOC/NRL, GMAO/NASA, JMA, NCEP, UKMO), a panel of experts on Numerical Weather Prediction and Air Quality forecast verification, a few data providers from EUMETSAT, NASA, and NOAA ESRL, and several observers from the research community (University of Leeds, Purdue University, LSCE, JPL).

NWP and AQ verification experts gave an overview of the common practices in operational forecast verification. They also outlined key questions to address in setting up a verification system. Modeling centers gave overviews related to their centers' current forecasting status and verification activities. Data providers gave updates on GALION

activities, on the status of the CALIPSO Level 1.5 NRT product, and on the SEVIRI aerosol product from EUMETSAT.

The meteorological operational community has approximately thirty years of experience in designing and implementing meaningful verification measures to check consistency and quality of the forecasts. A range of scores is used depending on the forecasting system (deterministic versus probabilistic) and on the verified variables. The aerosol forecasting community is now facing similar questions in regards to defining a set of scores against which the relative improvements of the forecasts can be measured. In recent years, there has been a rapid development of aerosol forecasting activities, and it is now time to discuss where this field is going. Many tools and know-how can be adopted from the meteorological community. However some specific aspects of aerosol verification need also to be considered. When a choice of scoring measures is made, this informs the line of future system assessments and research focus. The classical example in the NWP community is the 500hPa anomaly correlation coefficient, which is the correlation between the forecast and analyzed deviations from a chosen climate state. Most NWP centers use this score to gauge forecast skill improvements over time. There is a need to consider whether there is an equivalent to this measure that is needed for aerosol forecasting, or whether it is necessary to define something different. Regardless of the chosen scores, a reference climatology possibly based on observations would be useful for standardization of aerosol forecast verification.

At the same time, the current forecasting systems have developed a quasi-total reliance on NRT AERONET data for observation-based verification. It is in the interest of this community to explore the use of other ground-based quality aerosol datasets, such as those provided by the GAW stations. Other avenues to explore include the use of less conventional observations, such as profiles from the CALIOP lidar on the CALIPSO spacecraft or the ground-based lidar observations from MPLNET and AD-Net network maintained by NASA and NIES respectively. Gaining support for these types of observational activities from the big data agencies is a crucial step into ensuring the continuation and the improvement of the aerosol verification activities.

The purpose of the meeting was to initiate the discussion on a set of common verification measures that can be accepted as “standard” in the community and to understand what can be learned from the experience of the NWP and AQ communities. The key issues highlighted in the presentations and the discussion were those related to the definition of metrics that are representative of users’ needs, and that at the same time can be used for administrative purposes to show progress in the forecast skills to the funding agencies. Specific subject areas are covered below:

Why have a verification system and what are its attributes?

Four main reasons were identified:

- Monitor performance (administrative)
- Identify and correct model flaws for forecast improvement (scientific)
- Improve decision making and policies (economic)

- Understand biases and strengths/weaknesses of models (strategic)

The components of a well-designed verification system were also reviewed: (i) forecast attributes; (ii) observations availability/analysis; (iii) visualization; and (iv) reference system. For the aerosol forecasting community, there is still the need to identify a reference system (consensus climatology) and to ensure the availability of the verifying data over time. It was pointed out, for example, that it will become necessary to ask the data providers to maintain at least 10-12 ground measuring stations for the foreseeable future to ensure availability of a stable verifying data-set. The consensus climatology should ideally be based on model satellite and ground-based stations for the total/fine Aerosol Optical Depth (AOD) and on a combination of model and ground-based stations for Particulate Matter (PM₁₀/PM_{2.5}).

Overview of NWP and AQ verification measures

Some classic NWP scores were presented and discussed, such as bias, absolute error, RMS error, and the anomaly correlation (which is more suitable to identify matching dynamical patterns and is hence better for continuous variables). Scores based on a contingency table, such as frequency bias, hit rate, false alarm rate, equitable threat score, true skill score, can all be defined to check the skill of the aerosol forecasting system in predicting exceedances of given thresholds and/or extreme events. This is deemed useful in decision making and implementation of air quality policies. The use of confidence intervals was also greatly recommended.

Specifically from the AQ community, other useful scores were discussed, such as normalized bias/RMSE and fractional gross error (more resilient to outliers than standard bias/RMSE, but still fair both in case of under and over prediction), maps of fields with observed values plotted over model fields, several scores based on the contingency table, time averages of the scores, and scores by area type (i.e. rural sites, versus urban sites). Also recent developments in spatial verification were pointed out as possible avenues for meaningful aerosol verification.

There is a substantial network of surface AQ measurements across much of Europe and United States. Many of these measurements are available in NRT, but the quality control is not very stringent, the precision is often low, and the particulate matter is not speciated.

Ideally, a verification system for AQ applications should:

1. Provide metrics that evaluate standard field statistics and exceedance skill
2. Stratify evaluations according to site type and examine speciated components
3. Give a baseline comparison provided by persistence forecast
4. Benefit from innovative visualization (EPSgrams for the multi model ensembles, Taylor plots, Soccer plots showing bias versus total error, etc.)

Current status of aerosol forecast verification at operational centers

For most centers (ECMWF, NRL/FNMOC, GMAO) the verification is based mainly on AOD from AERONET data (fine/total). Measures computed routinely are: maps of bias and RMSE averaged over a month, time series of bias and RMSE, and single-station time series to monitor model performance over some regions of interest (i.e., Sahara desert, biomass burning regions in South America and Africa, Europe, etc.). The bias and RMSE are also computed as 24h means as a function of forecast range to increase representativeness of the sample and study the forecast behavior. In addition to ground-based stations, some centers use satellite data either routinely (AVHRR over selected regions at NRL) or experimentally (CALIOP aerosol backscatter data at ECMWF).

Dust verification is performed at NRL/FNMOC using visibility observations. The visibility observations are available on the WIS and are also used by meteorological forecasters. Some data, however, are not reliable due to several reasons: methodology of reporting, human errors, etc. A list of “good stations” according to a set of stringent criteria has been compiled by NRL and can be made available to the interested ICAP participants.

JMA presented a verification of the dust assimilation/forecast system based on WMO dust reports, gridded to model resolution and used to build a contingency table. One main point raised was the sensitivity of the scores to the threshold used to define a certain dust event.

NCEP computes many verification measures for its AQ forecasts of PM using NRT AIRNow and GOES Aerosol Smoke Product to verify smoke predictions. Verification is based on the accuracy with respect to the PM standard (currently $35 \mu\text{g m}^{-3}$ in the USA). Retrospective verification is also performed using aerosol composition observations from the STN and IMPROVE networks. Other variables such as PBL height observed from several sensors are also monitored.

UKMO highlighted the use of satellite TOA OLR observations as a tool for verification of global dust forecasts. Regional dust forecasts for Southern Asia are verified using both AERONET data and a NRT dust satellite product developed in-house at the Met Office, based on SEVIRI infrared brightness temperatures. Errors on the SEVIRI products are assessed from comparison with available AERONET stations. Spatial verification, using the SEVIRI product, is routinely conducted to monitor the model skill in producing realistic dust patterns, and to understand the different error sources (location, intensity of the dust storms, among others).

Some suggested scores

There was a consensus to define an Equitable Threat score event based using AERONET (fine/total AOD) and other observing networks (PM 2.5). A score for visibility was also deemed necessary.

After the workshop, a proposal was made to have a continuous score for AOD that would be the equivalent of the anomaly correlation for aerosols. This would be based on a consensus aerosol climatology, integrated from the surface upward and to provide the vertical integrated aerosol mass (or optical depth) between surface and specified pressure levels (for example at 1000, 925, 850 hPa). The anomaly correlation of such an aerosol quantity could define the quality of the aerosol between the surface and the relevant level in a way similar to the standard geopotential anomaly correlation. However, some believe that this type of score cannot work for aerosols fields, which have shorter correlation lengths than weather patterns. An anomaly correlation score for aerosols might end up highlighting dust events in Africa and Asia at the expense of more isolated urban pollution. Given also that the different centers have a markedly different customer base, such a uniform metric might be problematic.

Verification against own analysis was discussed briefly and seen as useful for model development and to monitor the performance of the forecast. Parallel verification of the underlying meteorology was also suggested, since variables such as surface temperature, surface winds, boundary layer height, humidity, precipitation, cloud fraction, TOA radiative fluxes, all play a role in the aerosol prediction. The importance of verifying emissions was also pointed out as so much of the forecast error in aerosol fields comes from the specification of the sources.

Data Availability for NRT verification

AERONET remains the most reliable source for AOD data. The potential of the GAW stations was mentioned.

NASA LaRC presented the level 1.5 CALIPSO product designed for NRT verification and data assimilation. Cloud-cleared, average attenuated backscattering (median and standard deviation) plus feature mask at 20km horizontal and 60 m vertical resolution will be provided operationally starting in 2011. Innovative aerosol scores can be computed from the unconventional lidar data, based on matching surface/near-surface aerosol extinction; mean upper troposphere extinction; mean-height (i.e. height where half of AOD is above, half below; or: 66% below, 95% below, etc.); aerosol scale height; scale height of best-fit exponential or more sophisticated functions, to mention a few.

Other possible sources of lidar data for verification are the lidar networks coordinated under the WMO GALION program. Some of the GALION networks (MPLNET, AD-Net) already provide NRT data. Other network such as EARLINET offer more sophisticated instrumentation but more limited capability for continuous measurements (4 sites at the moment). NRT developments are possible in the future, although the level 1.5 data would not be fully quality-assured. Limited funding resources are often the reason for the lack of NRT support.

As far as satellite datasets, MISR was recommended for verification purposes especially over land, having low bias and being well documented. However, MISR tends to underestimate mid-visible AOD for AOD above ~ 0.5 , especially over land. There are also several AOD products based on the SEVIRI (MSG) radiance observations that could provide NRT data over Europe and Africa, at high temporal resolution, but there is little political will on the part of the big data players to make those products operationally available.

Foreseeable collaborations on verification issues

It was suggested that all centers carry out a user survey to identify the specific needs that may inform the type of model skill that are more valuable. The results of this survey would be shared among the ICAP members and would provide motivations and guidance on the choices of verification scores.

Data sharing was discussed along with model intercomparison: both are seen as strengths. A suggestion was made to use an existing “template” from one of the centers and see if things can be ported to the other centers or developed within a common framework. The use of online tools (NCAR MET, for example) was also suggested.

Multi-model ensemble aerosol forecasting

The group discussed the potential to develop a multi-model ensemble for aerosol forecasting and the development of a baseline climatology for verification purposes. Most of the ICAP models include at least a prognostic dust species. For this reason, an action was put to the participating centers to look into sharing data and producing global maps of dust AOD using a common color bar. This effort is currently ongoing and some preliminary results will be presented during the upcoming third meeting of the ICAP members, which will focus on ensemble prediction and assimilation systems for aerosols.

Recent examples of multi-model ensembles in the AQ community were also reviewed, particularly the European multi-model ensemble system for Regional Air Quality prediction developed under the GEMS project. The implementation of this effort allowed for the development of useful tools for modelers to track model performance and spot problems; the coordination of different models within a unified framework, highlighting the benefits of common formats; and the collection of air quality data from various sources into one database, even with the expiration conditions attached to data usage by some providers. An added benefit of model inter-comparison is its utility to forecasters, who can learn about strengths and weakness of different models and make more informed forecasts. A framework similar to the European RAQ system could be developed for the global aerosol forecasting systems.

Conclusions

The meeting concluded with a list of topics for future discussions related to general issues facing the global aerosol forecasting community such as ensemble systems, direct radiance assimilation, model variable requirements, product development and delivery, and emission datasets. Participants were keen on continuing to collaborate and communicate over these topics of common interest. The next meeting will be held in Boulder, CO, in May 2011 and will discuss ensemble forecast and assimilation systems for aerosol prediction.

AD-Net	Asian Dust Network
AERONET	Aerosol Robotic Network
AOD	Aerosol Optical Depth
AQ	Air Quality
AVHRR	Advanced Very High Resolution Radiometer
CALIOP	Cloud-Aerosol Lidar with Orthogonal Polarization
CALIPSO	Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observations
EARLINET	European Aerosol Research Lidar NETwork
ECMWF	European Centre for Medium-Range Weather Forecasts
ESRL	Earth System Research Laboratory
EUMETSAT	European Organization for the Exploitation of Meteorological Satellites
FNMOCC	Fleet Numerical Meteorology and Oceanography Center
GALION	GAWAerosol Lidar Observation Network
GAW	Global Atmospheric Watch
GEMS	Global Earth-system Monitoring using Space and in-situ data
GMAO	Global Modeling and Assimilation Office
GOES	Geostationary Operational Environmental Satellite
ICAP	International Cooperative for Aerosol Prediction
IMPROVE	Interagency Monitoring of Protected Visual Environments
JAXA	Japanese Aerospace Exploration Agency
JMA	Japan Meteorological Agency
LSCE	Laboratoire des Sciences du Climat et l'Environnement
MET	Model Evaluation Tools
MISR	Multi-angle Imaging SpectroRadiometer
MODIS	Moderate Resolution Imaging Spectroradiometer
MPLNET	Micropulse Lidar Network
MSG	Meteosat Second Generation
NASA	National Aeronautical and Space Administration
NCAR	National Center for Atmospheric Research
NCEP	National Centers for Environmental Prediction
NIES	National Institute for Environmental Studies
NOAA	National Oceanic and Atmospheric Administration
NPP	NPOESS Preparatory Project
NRL	Naval Research Laboratory
NRT	Near Real Time
NWP	Numerical Weather Prediction
PBL	Planetary Boundary Layer
OLR	Outgoing Longwave Radiation
RAQ	Regional Air Quality
RMS	Root Mean Square
SEVIRI	Spinning Enhanced Visible and Infrared Imager
STN	Speciation Trends Network
TOA	Top Of the Atmosphere
UKMO	United Kingdom Meteorological Office

WIS	WMO Information System
-----	------------------------